Analysis of Traffic Distribution in Cellular Networks

P J Smith[#], A Sathyendran[*] and A R Murch[*]

- Victoria University of Wellington, New Zealand, * - Vodafone New Zealand Ltd.

Vodafone New Zealand Ltd., Level 6, 21 Pitt Street, Private Bag 92161,

Auckland, New Zealand Ph. No. +64 9 357 0326, Fax No. +64 9 379 3451,

Email : arasaratnam.sathyendran@vodafone.co.nz

Abstract

Accurate air interface traffic forecasting and dimensioning is of importance in any cellular network for achieving cost and quality requirements. A previous paper [1] analysed the appropriateness of the Erlang B model to estimate the mean call blocking experienced by cellular traffic using the traditional confidence interval method. This paper presents a modified confidence interval method to compare the mean blocking of the measured data and Erlang B results. In addition to a more complete study of the mean, the blocking distribution is also considered. The Erlang Loss Model (ELM) is studied to completely characterise the distribution of blocking using the ELM. Exact expressions for the busy time distribution are derived for this study.

The results presented in this paper indicate that Erlang B formula is an appropriate model for calculating mean call blocking on the air interface. The ELM on the other hand appears to be rather a poor model for the overall blocking distribution. Further study is needed to establish the appropriateness of Erlang B formula as a general tool.

I. Introduction

Radio resource allocation is a critical part of cellular network planning, as there is only a limited amount of radio spectrum available for cellular use. Radio resource allocation is based on the traffic carried by the cells and traffic tables are used to aid the allocation. Generally Erlang B traffic tables have been used for dimensioning cellular networks. However, due to the nature of cellular traffic and features that are implemented to improve cellular network capacity, cellular traffic violates a number of Erlang B assumptions. Incorrect traffic thresholds can either result in unacceptable call blocking (and customer dissatisfaction) or excessive infrastructure cost (and wastage of precious radio resource). Thus, it is crucial to establish accurate traffic thresholds to dimension cellular networks.

Claims have been made about the appropriateness or otherwise of Erlang B to model call blocking on the air interface [2]. However, few studies have been published detailing how appropriate Erlang B is in modeling call blocking on the air interface [3-5]. Generally the validity of the Erlang B distribution to model cellular network traffic has been questioned in these studies. Improved and usually complicated models have been presented to estimate the blocking performance. From a network operators point of view it is difficult to derive blocking models based on the new distributions suggested in many of these studies. A previous paper [1] analysed the appropriateness of the Erlang B model to estimate the mean call blocking experienced by cellular traffic. The traditional confidence interval method was used to determine the appropriateness of Erlang B model for dimensioning cellular network. The traditional confidence interval method assumes that the data is normally distributed or appeals to the central limit theorem for its validity. However, the measured data is highly non-normal. Not only is it highly skewed but also the data stems from a mixed distribution where there is a non-zero discrete probability of zero blocking and the rest of the probability is continuously spread.

In this paper a modified confidence interval method is presented to compare the mean blocking of the measured data and the Erlang B result. A transformation to force proportions into approximate normality is used and confidence intervals are produced in the transformed space. Results from this approach act as a sanity check on the previous results presented in [1].

In addition to a more comprehensive study of mean, blocking distribution is also considered and the distribution of blocking is completely characterised by studying the ELM (the finite time version of the Erlang B result). The reason for this is as follows. The Erlang B blocking probability is a steady state result and hence gives no indication about how likely large blocking probabilities are. Such information is useful in planning and the question is can the distribution of the blocking time in the busy hour is derived? In the literature the behaviour of this model has been studied for finite time periods via Laplace transforms [6] and more recently by large deviations methods [7]. However, to the best of the authors' knowledge the distribution of the time spent in a given state has not been derived. Hence in this paper exact expressions for the busy time distribution are derived. This has the added advantage of requiring no inversion of Laplace transforms.

This paper is organised as follows. In Section II the modified confidence interval model is presented and the results obtained

0-7803-5565-2/99/$10.00 © 1999 IEEE

using this model are compared with the Erlang B results. In Section III the blocking distribution of the measured data is compared with that of the ELM. Section IV gives the conclusions. The Appendix describes the method used to derive the busy time blocking distribution predicted by the ELM.

II. Comparison of Mean

The measured data used in this paper was gathered over a period of one year from Vodafone New Zealand's network. Each measurement point represents a measurement from the busy hour. The offered traffic and blocking probability was estimated from the busy hour carried traffic and busy hour congestion times. The data was obtained for cells with 7, 14 and 21 channels. Figure 1 shows the distribution of blocking for an offered traffic of 1.5 Erlangs for a 7 channel cell.

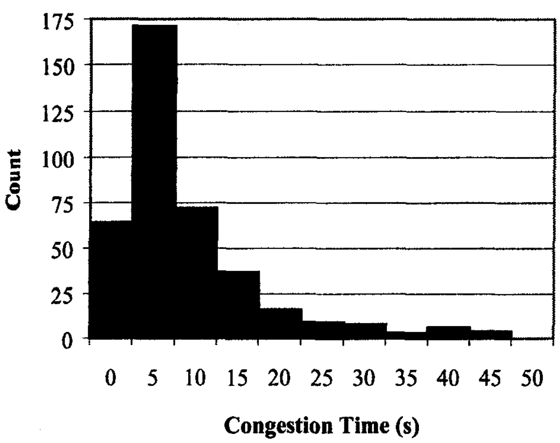

Figure 1 : Distribution of measured blocking for an offered traffic of 1.5 Erlangs for a 7 channel cell.

From Figure 1 it can be seen that the blocking distribution is highly non-normal. Thus, it may not be appropriate to use the standard confidence interval method to determine the appropriateness of Erlang B model for dimensioning cellular networks. In this paper a modified confidence interval method is used. The following transformation is made to transform the data [8]

$$y_i = \ln\left(\frac{x_i}{1-x_i}\right) \qquad \text{...(1)}$$

where x_i is the original measured blocking probability data, y_i is the transformed data. However, when x_i is zero eqn. (1) is undefined. Thus probability of zero blocking is handled separately by adjusting the lower limit of the confidence interval. The transformed data is shown in Figure 2.

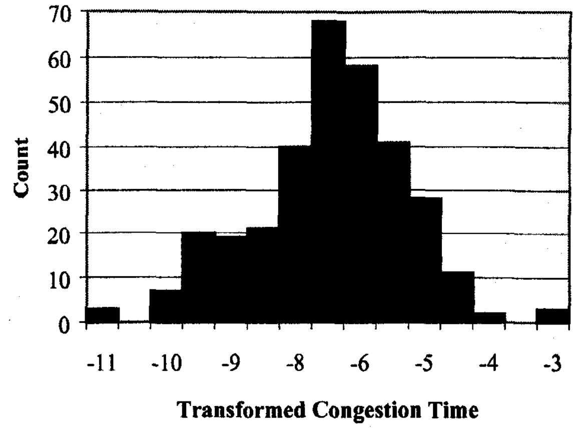

Figure 2 : Distribution of transformed blocking data for an offered traffic of 1.5 Erlangs for a 7 channel cell.

From Figure 2 it can be seen that the transformed data is closer to normal distribution than the data presented in Figure 1.

To compare the measured data with Erlang B, the measured data is divided into 0.2 Erlang bins spaced at 0.5 Erlangs. The confidence interval for E(Y) is computed using the standard confidence interval formula

$$\bar{y} \pm \Gamma \times \frac{\sigma_y}{\sqrt{n}}, \qquad \text{...(2)}$$

where n is the number of samples in the bin, σ_y is the standard deviation of the transformed data in the bin considered, Γ is the desired level of significance and $\bar{y}$ is the mean of the transformed data in the bin. The confidence interval is in the transformed space, i.e. in $\ln\left(\frac{x_i}{1-x_i}\right)$ form.

An approximate confidence interval for E(X) is obtained by taking the inverse transform of the interval for E(Y) as below

$$\left(\frac{\exp(L)}{1+\exp(L)}, \frac{\exp(U)}{1+\exp(U)}\right) \qquad \text{...(3)}$$

where L and U are the lower and upper limits calculated using eqn. (2). Figure 3 shows the mean blocking results and confidence intervals produced from the methods described above. These results are compared with mean blocking figures calculated using the Erlang B formula.

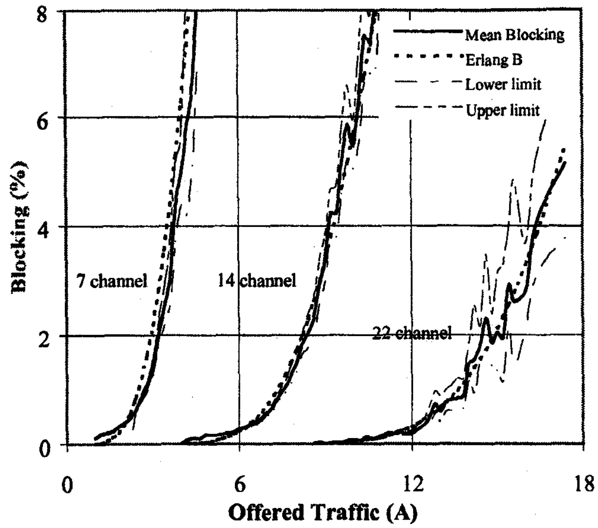

Figure 3: Offered traffic versus blocking plotted for the measured data and Erlang B.

The results in Figure 3 indicate that the Erlang B blocking curve falls within the measured data confidence interval more often when there are more channels (greater than 14) per site and when blocking is above 1%. Conversely Figure 3 shows that Erlang B falls tend to fall outside the measured data confidence interval when there is fewer channels per site (7 channels) and when blocking is less than 6%. Where Erlang B is less appropriate (i.e. channels less than about 14) Erlang B over estimates call blocking (i.e. it suggests the blocking is worse than it really is). This indicates that cellular traffic in cells with a small number of channels is smoother than assumed by Erlang B. Nevertheless the results indicate that Erlang B is a good approximation for the mean blocking for most instances, especially when the simplicity of the Erlang B formula is considered.

The similarity in the conclusions derived by the method described here and those derived in [1], is probably due to the large number of sample points in each bin. Due to these large sample sizes the central limit theorem seems to yield valid results. This paper provided a check on the previous results.

In the following section the distribution of the measured data is compared with that derived under Erlang B assumptions (the ELM). This provides further evidence as to whether Erlang B type assumptions provide a good model for cellular traffic in terms of the whole distribution as well as the mean.

III. Blocking Distribution

In this section the distribution of blocking of the measured data is compared with that derived under Erlang B assumptions. Since the Erlang B model appears to be satisfactory for the mean blocking in the busy hour [1] the standard ELM [7] is used to model the traffic. The behaviour of this model has been studied previously [6,7]. However, an analytic expression for the finite time blocking distribution appears to be unavailable. Hence exact expressions for the busy time distribution are derived which are, with one exception, a direct formulation requiring only Poisson and binomial terms. Refer the Appendix for the detailed derivation of the equations.

Figure 4 compares the CDFs of the measured and analytically calculated CDFs for 7 channel cells for different offered traffic. The label "Empirical" refers to results obtained from the measured data. The label "Model" refers to results obtained from the model derived in this paper. The numbers at the end of the labels "Empirical" and "Model" refer to the offered traffic in Erlangs (A). The offered traffic figures 2.51, 2.93 and 3.52 correspond to mean blocking of 1%, 2% and 4% respectively, in a 7 channel cell.

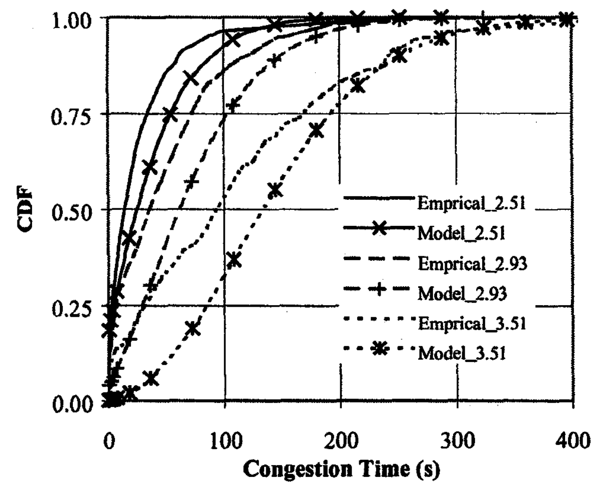

Figure 4 : Measured and ELM distributions verses congestion time for a 7 channel cell.

The results show that the measured blocking distributions do not match the analytical model very well. Certainly the match is much worse than the mean comparison which were quite satisfactory. In particular at low blocking values (say 1% or less) the model predicts a higher blocking probability of zero blocking than that observed and at higher blocking (say >2%) the model predicts less than that observed. There may be a number of reasons why the Erlang B distribution is a good approximation to the mean when the loss model gives poor approximation to the distribution. These are tabulated in Table 1.

It has been concluded that there is a lack of fit between the measured data and the ELM based on the results presented for a 7 channel cell. It could be argued that the mismatch between mean Erlang B blocking and measured data for a 7 channel cell (refer Figure 3), contributed to the lack of fit in Figure 4. However, the authors feel that similar conclusions could be derived for 14 and 22 channel cells. Further study is required to prove this.

Table 1 : Reasons for lack of fit between the empirical data and ELM.

Model Assumption	Reasons for ELM assumption may be flawed
Poisson arrivals at rate λ	1. Different cells may have different λ's 2. Traffic may not be constant in the hour
Exponential call times, rate μ	1. Different cells may have different, μ's 2. Call length distributions may be non-exponential. Short calls are more likely than the exponential gives due to handover and ping ponging of mobile at the cell boundary. Longer calls are less likely than the exponential gives due to handover.
Blocked calls leave the system	Blocked calls are reattempted in adjacent cells.
Infinite population	Only those in the coverage area of the cell can connect to that cell
λ and μ are known	In practice μ is estimated from the mean call length and λ is gained from $\rho = \lambda / \mu$.

It is difficult to identify which are the most likely reasons. One possibility may be the error in λ and μ values. These figures were estimated from the mean call length and not directly measured. Since the mean call length differs between cells, a range of values has been approximated by a single pair of constants. Also the shape of the distribution is dependent on the mean call length used. In the estimation of results a mean call length of 90 seconds was used. In this case for an offered traffic of 2 Erlangs and a 7 channel cell, the model gives a probability of 0.528 for zero blocking whereas the figure for measured data is 0.04. If a mean call length of 120 seconds is used, then the zero blocking probability estimated from the model is 0.385. Hence, variations in the mean call length can cause the curves to become more or less similar. Another possible reason is that the Erlang B formula does not require an exponential service time. Hence, if the real service distribution is short tailed, compared to exponential, but with same mean, then the Erlang B results for the mean blocking is still valid but results presented in this paper on the blocking distribution are not.

To eliminate the error in the estimation of λ and μ values the blocking distribution calculated form the model has to be compared with measured data from individual cell. This would enable to get an accurate estimate of λ and μ values and thus provide a better comparison of the model with measured data. Further study is need in this area.

IV. Conclusions

From the results presented in this paper Erlang B formula appears to be an appropriate model for calculating mean call blocking on the air interface when the number of channels per cell site is greater than about 14 and call blocking greater than about 1%. Where Erlang B is less appropriate (i.e. channels less than about 14 and blocking less than 4%) it over estimates call blocking (i.e. it suggests blocking is worse than it really is). This indicates that cellular traffic in cells with a small number of channels is smoother than assumed by Erlang B. Nevertheless for all situations the Erlang B results are highly satisfactory for such simple model.

The ELM on the other hand appears to be rather a poor model for the overall blocking distribution. Since the assumptions behind the ELM are the same as those behind the Erlang B results, this casts some doubt on the use of the Erlang B formula as a general tool. In summary the assumptions are giving good matching to the mean but bad matching to the distribution.

Further study has to be done to compare the blocking distribution with measured data obtained from individual cells with accurate estimate of λ and μ values.

Appendix

Derivation of Busy Time Distribution

The derivation of the distribution of blocking under Erlang B assumption is described in this Appendix.

Consider the Erlang loss model with M channels, Poisson call arrival (rate λ) and exponential service time (rate μ). Let $X(\tau)$ be the continuous time Markov chain defined by $X(\tau)$ equals the number of channels in use at time τ for $\tau \in [0, t]$. Also define $S_i(t)$ to be the total busy time (time in state M) in [0,t] given the initial state X(0)=i. The unconditional busy time is denoted by S(t) and its distribution is characterised by the distribution function

$$F(x) = P(S(t) \le x) = \sum_{i=0}^{M} P(S_i(t) \le x) P(X(0) = i) \qquad ...(4)$$

If we assume that the system is in steady state at $\tau = 0$ then the initial state probability is given by the Erlang B or truncated Poisson formula [9]

$$P(X(0) = i) = q_i(\lambda) \left[\sum_{j=0}^{M} q_j(\lambda) \right]^{-1}, \qquad ...(5)$$

where $q_i(\lambda) = \lambda^i \exp(-\lambda) / i!$. The key to computing the first probability in eqn. (4) is to use a technique known as uniformization, which makes all states have the same rate of transitions by introducing intra-state transitions. The new process is completely equivalent to the old one but has different transition probabilities given by [9]

$$\rho^*_{ij} = \begin{cases} 1 - v_i/v & j = i \\ (v_i/v)p_{ij} & j \neq i \end{cases} \quad \text{...(6)}$$

where ρ_{ij} are the transition probabilities of the original process, v_i is the rate the original process leaves state i and $v = \max(v_0,\dots,v_m)$ is the common rate at which the new process leaves any state. For a birth and death process like the ELM these terms are well known [9] and are listed below

$$v_i = \begin{cases} \lambda + i\mu & i = 0,1,\dots,M-1 \\ m\mu & i = M \end{cases} \quad \text{...(7)}$$

$$v = \lambda + (m-1)\mu, \qquad \text{assuming } \lambda > \mu \quad \text{...(8)}$$

$$\begin{aligned} &\rho_{i,i+1} = \lambda(\lambda + i\mu)^{-1} = 1 - \rho_{i,i-1}, \qquad i = 1,2,\dots,M-1 \\ &\rho_{0,1} = \rho_{1,0} = 1 \end{aligned} \quad \text{...(9)}$$

Using the uniformized process we can compute the first probability in eqn. (4) by conditioning on N(t), the number of transitions in [0,t] made by the new process, and noting that N(t) is a Poisson process with rate v, hence [9]

$$P(S_i(t) \le x) = \sum_{n=0}^{\infty} P(S_i(t) \le x | N(t) = n) q_n(vt). \quad \text{...(10)}$$

Now defining Z(t) to be the total number of visits to the busy state in [0,t] and X(k) to be the total time spent in the busy state during k visits we have

$$P(S_i(t) \le x) = \sum_{n=0}^{\infty} q_n(vt) \sum_{k=0}^{n+1} \begin{matrix} P(Z(t) = k | X(0) = i, N(t) = n) \\ \times P(X(k) \le x) \end{matrix} \quad \text{....(11)}$$

From Ross [9] the last probability in eqn. (11) is the binomial tail probability

$$P(X(k) \le x) = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{x}{t}\right)^i \left(1 - \frac{x}{t}\right)^{n-i}. \quad \text{...(12)}$$

The conditional probability for Z(t) in eqn. (11) is harder and closed form expressions seem infeasible for anything other than the smallest cases, M=2,3 say. Hence we use first step analysis [10] to create a recursion. Denoting

$$q(k,i,n) = P(Z(t) = k | X(0) = i, N(t) = n) \quad \text{...(13)}$$

for compactness, this gives

$$\begin{aligned} q(k,i,n) &= \sum_{j=i-1}^{i+1} p^*_{ij} q(k,j,n-1) \qquad i = 0,1,\dots\dots M-1 \\ q(k,M,n) &= \sum_{j=M-1}^{i+1} p^*_{Mj} q(k-1,j,n-1) \end{aligned} \quad \text{...(14)}$$

which can be used as a recursion to the boundary values

$$\begin{aligned} &q(0,M,0) = 0 \\ &q(0,i,0) = 1, \qquad i = 0,1,\dots,M-1 \\ &q(n+1,M,n) = \left(p^*_{MM}\right)^n \\ &q(n+1,i,n) = 0, \quad i = 0,1,\dots\dots,M-1 \end{aligned} \quad \text{...(15)}$$

Hence the distribution of the busy time is given exactly by eqns. (4) – (15).

Acknowledgment

The authors would like to thank Giselle Baker for providing the measurement data and Vodafone New Zealand for allowing us to make use of the network data..

References

[1] G W Tunnicliffe, A R Murch, A Sathyendran and P J Smith "Analysis of traffic distribution in cellular networks," VTC Proceedings 1998, Ottawa, Canada.

[2] William C.Y. Lee " Mobile Cellular Telecommunications Systems " McGraw Hill International Editions, 1989.

[3] Chris Jedrzycki and Victor C.M.Leung "Probability distribution of channel holding time in cellular telephony systems " *Proceedings of IEEE 44th VTC*, pp. 247-251, Sweden 1994.

[4] David E. Everitt, "Traffic engineering of the radio interface for cellular mobile networks " *Proceedings of IEEE.*. VOL 82, NO 9. pp. 1371-1382, September 1994.

[5] S H Bakry and MH Ackroyd, "Teletraffic analysis for single-cell mobile radio telephone systems," *IEEE Trans. Commun.*, vol. COM-29, pp. 298-304, March 1981.

[6] V E Benes, "Mathematical Theory of Connecting Networks and Telephone Traffic," Academic Press, New York, 1965.

[7] A Schwartz and A Weiss, "Large Deviations for Performance Analysis : Queues, Communications and Computing," Chapman and Hall, London, 1995.

[8] N R Draper and H Smith, "Applied Regression Analysis," Wiley and Sons, Inc., New York, 1981.

[9] Sheldon Ross, "Stochastic Process," John Wiley & Sons, Inc., New York, 1996.

[10] H M Taylor and S Karlin, "An Introduction to Stochastic Modelling," Academic Press, Inc., Orlanda 1984.